Putative Alleles for Increased Yield from Soybean Plant Introductions

E. A. Kabelka,* B. W. Diers, W. R. Fehr, A. R. LeRoy, I. C. Baianu, T. You, D. J. Neece, and R. L. Nelson

ABSTRACT

Improving seed yield of soybean [*Glycine max* (L.) Merr.] is an important breeding goal. The objective of this study was to evaluate two soybean PIs as sources of alleles for the enhancement of seed yield in North American cultivars. A population of 167 F_5–derived lines was developed from a cross between 'BSR 101' and the experimental line LG82-8379. BSR 101 has nine of 10 major ancestral lines contributing to the commercial gene pool of North America, while LG82-8379 was selected from a cross between PI 68508 and FC 04007B. The F_5–derived lines, divided into three sets based on maturity, were evaluated for 145 polymorphic simple sequence repeat (SSR) marker loci and for seed yield and other agronomic traits in 12 environments. Fifteen quantitative trait loci (QTL) were significantly [$P < 0.05$, likelihood of odds (LOD) > 2.5] associated with seed yield in at least one set with two significant across all sets. For nine of the yield QTL, the LG82-8379 alleles were associated with yield increases of 1.7 to 5.4% while the BSR 101 alleles increased yield 2.4 to 4.4% at six yield QTL. Four yield QTL were associated with significant changes in R8, eight with plant height, and three with seed protein concentration. Additional QTL were identified for R8, plant height, lodging, and seed protein and oil concentration. These results indicate that soybean PIs have the genetic potential for improving seed yield of U.S. soybean cultivars.

Improving seed yield in soybean is an important goal of breeding programs. In North America, soybean yield improvement comes mainly from selection in populations from crosses between commercial cultivars or experimental lines derived from these cultivars. Pedigree and genetic marker studies have shown that the genetic base of the North American soybean gene pool is narrow (Apuya et al., 1988; Keim et al., 1989; Gizlice et al., 1993, 1994; Thompson et al., 1998). This narrow genetic base has occurred primarily from using a small number of ancestral lines as the founding stock (Gizlice et al., 1993, 1994).

Introgressing exotic germplasm into cultivars to increase genetic diversity within domesticated crops has been used to enhance complex traits such as yield (Tanksley and McCouch, 1997). In soybean, Thompson and Nelson (1998) tested experimental lines derived from crossing North American cultivars with several plant introductions. Lines that yielded significantly more than their North American parent were identified which indicated that soybean PIs may have alleles for enhanced seed yield.

Molecular marker analysis can aid in the discovery and mapping of QTL associated with beneficial and novel alleles from exotic parents. For example, in soybean, two populations of recombinant inbred lines (RILs) derived from 'Minsoy' and 'Archer', and 'Noir 1' and Archer (Orf et al., 1999) have been used extensively for the identification of QTL governing agronomic and seed composition traits utilizing molecular markers (Orf et al., 1999; Terry et al., 2000; VanToai et al., 2001). In one study comparing these two RIL populations, Orf et al. (1999) identified QTL for plant height, lodging, days to flowering and maturity, reproductive period, seed yield, seed weight, and seed oil and protein concentration. At many of the QTL identified in this study, the beneficial or novel alleles were those provided by Minsoy and Noir 1. Minsoy and Noir 1 are divergent in lineage from each other and from North American cultivars (Mansur et al., 1993a, 1993b; Orf et al., 1999), whereas Archer is a North American cultivar (Cianzio et al., 1991).

Yield alleles from exotic sources are not well documented in soybean and those identified have often been related to other agronomic traits such as maturity. In this study, a soybean population consisting of F_5–derived lines was developed from a cross between 'BSR 101' and the experimental line LG82-8379. The pedigree of BSR 101 has nine of 10 ancestral lines that have made the greatest contribution to the commercially used gene pool of North America (Bernard et al., 1988; Gizlice et al., 1994) whereas LG82-8379 was selected from a cross between two soybean PIs, PI 68508 and FC 04007B. Using random amplified polymorphic DNA markers, the two PIs had been shown to be genetically distinct from all major ancestral lines of modern North American soybean cultivars (Thompson et al., 1998). From this cross, high-yielding and genetically diverse experimental lines have been developed (Thompson and Nelson, 1998) and released (Thompson et al., 1999). The objective of this study was to evaluate the two PIs as sources of alleles for the enhancement of seed yield in North American cultivars.

MATERIALS AND METHODS

A soybean population consisting of 167 indeterminate F_5–derived lines was developed by single seed descent from a cross between BSR 101, a maturity group (MG) II cultivar, and the MG IV experimental line LG82-8379. LG82-8379 was derived from a cross between two soybean PIs, PI 68508 and FC 04007B. Each line descended from a different F_2 seed

E.A. Kabelka and B.W. Diers, Dep. of Crop Sci., I.C. Baianu and T. You, Dep. of Food Sci. and Human Nutrition, and D.J. Neece and R.L. Nelson, USDA-ARS, Soybean/Maize Germplasm, Pathology, and Genetics Research Unit, Dep. of Crop Sci., Univ. of Illinois, Urbana, IL 61801; W.R. Fehr, Dep. of Agronomy, Iowa State Univ., Ames, IA 50011; A.R. LeRoy, Dep. of Agronomy, Purdue Univ., West LaFayette, IN 47907. Received 13 Dec. 2002. *Corresponding author (ekabelka@uiuc.edu).

Published in Crop Sci. 44:784–791 (2004).

677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: LOD, likelihood of odds; MG, maturity group; PCR, polymerase chain reaction; QTL, quantitative trait locus/loci; R1, days to flower; R8, days to maturity; RIL, recombinant inbred line; SSR, simple sequence repeat.

produced from one of three F_1 plants. The F_5–derived lines were evaluated at six locations in both 2000 ($F_{5:8}$) and 2001 ($F_{5:9}$). Field locations in each year were Arthur, Hume, Ivesdale, and Urbana, IL; West Lafayette, IN; and Ames, IA. A nested split-plot design with two replications was used at each location. Lines were grouped into three sets based on previously determined maturities. The sets were designated Set 1 (54 lines), Set 2 (55 lines), and Set 3 (58 lines). Lines within sets were randomly assigned to subplots. Within a replication, sets were randomly assigned to whole plots. Cultivars BSR 101, Loda (MG II), and IA2038 (MG II) were included as checks within Set 1, cultivars IA2038 and IA3010 (MG III) were included as checks within Set 2, and cultivars IA2038, IA3010, and HS93-4118 (MG IV) were included as checks within Set 3. LG91-7350R (MG IV), a high-yielding experimental line previously developed from the BSR 101 × LG82-8379 population (Thompson et al., 1999), also was included in Sets 2 and 3 as a check. The experimental line LG82-8379 is heterogeneous and was not included in the field trials since it does not represent the exact parental genotype used in creating the BSR 101 × LG82-8379 population. Subplots were four rows wide with 0.61 to 0.76 m row spacing depending on the location. Row lengths were 3 m and seed was sown at a rate of 33 to 41 seeds m^{-1} row^{-1}. Subplots were not end-trimmed prior to harvest. Conventional tillage practices were followed at all locations maintaining a weed-free environment and recommended fertilization levels were applied.

Days to maturity, plant height, lodging, and seed yield were recorded in all trials. Days to maturity was recorded as the number of days after planting when approximately 95% of the pods had reached mature pod color (R8; Fehr et al., 1971). Plant height (mm) was measured at maturity as the average distance from soil surface to the apex of the main stem. Lodging was scored at maturity on a scale of 1 to 5, with 1 designated as plants standing erect and 5 as plants prostrate. The two center rows of each subplot were harvested and seed yield, expressed as kg ha^{-1}, was adjusted to 130 g kg^{-1} moisture. At three locations each year, days to flower was recorded as the number of days after planting when approximately 50% of the plants had at least one flower (R1; Fehr et al., 1971). Reproductive period was calculated by subtracting R1 from R8. Seed protein and oil concentration were measured on whole beans from four locations each year using near infrared transmission (NIT) grain analyzers, models ZX800 and ZX880 (Zeltex Inc., Hagerstown, MD). All samples were measured in duplicates in order to improve reliability. The seed protein and oil concentrations are reported on a moisture-free basis as grams per kilogram.

Leaf tissue DNA was extracted from eight greenhouse-grown seedlings per $F_{5:8}$ line according to Keim et al. (1988) with modifications as described by Kisha et al. (1997). The SSR markers used in this study were developed by P.B. Cregan (USDA-ARS, Beltsville, MD). Nonlabeled and fluorescently-labeled primers were obtained from Applied Biosystems (Foster City, CA) and Research Genetics (Huntsville, AL). Polymerase chain reactions (PCRs) were performed according to Cregan and Quigley (1997). The nonlabeled PCR products were analyzed by electrophoresis in 6% nondenaturing polyacrylamide gels (Sambrook et al., 1989; Wang et al., 2003) and stained with 1 μg mL^{-1} ethidium bromide. The fluorescently-labeled PCR products were analyzed using an ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, CA).

More than 500 SSR markers covering the 20 chromosomes of the soybean genome were screened against the parent BSR 101 and a set of seven DNA bulks that each consisted of a mixture of 22 different $F_{5:8}$ lines to identify markers segregating within this population. The bulk DNA set of $F_{5:8}$ lines was used instead of LG82-8379 to identify polymorphisms because LG82-8379 is heterogeneous and the exact parental genotypes are unknown. A total of 145 polymorphic SSR markers were identified. For each polymorphic marker, only two alleles were present within the population and allelic frequencies at each locus were as expected within and across sets. Since only two alleles were found at each locus, this suggests that the three F_1 plants used to create this population were homogeneous at the loci evaluated in this study. To determine their contribution, PI 68508 and FC 04007B were each genotyped with SSR markers associated with yield enhancement.

Agronomic traits were subjected to ANOVA by the PROC GLM and VARCOMP functions of SAS (Statistical Analysis System version 8.0, SAS Institute, Cary, NC). Sets, based on maturity, were considered as a fixed effect whereas environment, replications within environment, lines within sets, environment by sets, and environment by lines within set interactions were considered as random effects. Each location–year combination was considered an environment in the analysis. Significance was determined by an F test as described by McIntosh (1983), with the numerator and denominator degrees of freedom approximated as described by Satterthwaite (1946). Means, ranges, and standard deviations across environments within each set were computed for each trait. Least significant differences between set means were determined at a 5% significance level. Broad-sense heritabilities of each trait were calculated on a line-mean basis according to Fehr (1987). Pearson correlation coefficients among all traits within and across sets were calculated from the means of lines across environments using PROC CORR in SAS. The experiment-wide error rate in this study was calculated according to Lander and Botstein (1989).

Single marker-trait and composite interval mapping methods were used to identify marker association with each agronomic trait. For both methods, sets were analyzed separately and combined. Single marker-trait analysis was performed by PROC GLM in SAS. Marker genotypes, used as class variables, were considered as fixed effects whereas environments, replications, and lines were considered random effects. In the combined analyses, means of all lines from the three sets were analyzed together to detect QTL. Pairwise comparisons of least square means were used to test the significance ($P < 0.05$) of each marker class. Lines heterozygous for the genetic markers were included in all analyses, but their means were not included in the summary tables. Composite interval mapping was performed using the multiple-regression based computer program PLABQTL (Utz and Melchinger, 1996), with the required genetic linkage map constructed with the computer program JoinMap v. 3.0 (Van Ooijen and Voorrips, 2001). A minimum LOD of 3.0 and a maximum distance of 50 cM was used for testing linkages among markers. A minimum LOD of 2.5 was chosen to declare a marker association significant. The R^2 value was used to describe the phenotypic variance explained by individual markers. The total phenotypic variance explained by two or more QTL for a given trait within a single set or across sets was determined by using a multifactor ANOVA.

RESULTS

Field Data Analysis

Significant ($P < 0.05$) differences were observed among sets for all traits except seed yield (Table 1). Days to flower, reproductive period, plant height, lodging, and protein concentration increased with maturity

Table 1. Means, ranges, standard deviations (SD), and broad-sense heritability (H^2) estimates based on means for traits measured in the BSR 101 × LG82-8379 soybean population of F_5–derived lines grouped within Sets 1, 2, and 3, according to maturity.

Set	Mean†	Range	SD	H^2	Set	Mean	Range	SD	H^2
		Days to flower (R1)					Lodging score (1–5)‡		
1	33	29–37	1.7	0.87	1	2.2	1.4–3.3	0.5	0.83
2	38	36–42	1.4	0.81	2	2.6	1.6–3.7	0.5	0.88
3	41	39–44	1.1	0.64	3	2.7	1.7–3.4	0.4	0.80
LSD (0.05)§	0.5				LSD (0.05)	0.2			
		Days to maturity (R8)					Seed protein concentration, g kg^{-1}		
1	110	106–117	2.4	0.90	1	424	397–437	8.3	0.77
2	124	118–127	2.1	0.92	2	425	399–441	6.7	0.68
3	130	124–135	1.7	0.89	3	428	407–446	6.7	0.70
LSD (0.05)	1.1				LSD (0.05)	2.9			
		Reproductive period (R8–R1)					Seed oil concentration, g kg^{-1}		
1	77	73–85	2.3	0.82	1	221	203–229	4.5	0.81
2	85	80–90	1.9	0.76	2	218	206–225	3.7	0.77
3	89	83–95	1.6	0.66	3	217	207–228	3.7	0.81
LSD (0.05)	0.9				LSD (0.05)	1.6			
		Plant height, mm					Seed yield, kg ha^{-1}		
1	850	720–960	48.0	0.88	1	2841	2231–3283	177	0.68
2	940	820–1060	61.0	0.77	2	2811	2479–3071	137	0.48
3	970	880–1050	40.0	0.67	3	2806	2483–3168	156	0.77
LSD (0.05)	1.9				LSD (0.05)	62			

† Days to maturity from planting, plant height, lodging, and seed yield averaged across 12 environments; days to flower from planting, and reproductive period averaged across six environments; and seed protein and oil concentration averaged across eight environments.
‡ 1 = all plants standing erect, 5 = all plants prostrate.
§ LSD, least significant difference among sets.

whereas oil concentration decreased with maturity. Significant variation among lines within each set was present for all traits (Table 1). Set 1 lines had the greatest range for R1 and R8. The latest maturing Set 1 line was 7 d earlier than the earliest Set 3 lines, but some Set 1 lines had reproductive periods longer than the Set 3 lines. Set 1 lines exhibited the lowest and the highest seed yield of this population, ranging from 2231 to 3283 kg ha^{-1}.

The average seed yield of lines within Set 1 was approximately 4% higher than the average seed yield of 2726 kg ha^{-1} for BSR 101, but BSR 101 matured 8 d earlier than the earliest lines in Set 1. The mean yield of the checks were 3286 kg ha^{-1} for Loda, 3319 kg ha^{-1} for IA2038, 3733 kg ha^{-1} for IA3010, 3759 kg ha^{-1} for HS93-4118, and 3155 kg ha^{-1} for LG91-7350R. While lines within sets yielded as high as the MG II cultivars Loda and IA2038, and the MG IV experimental line LG91-7350R, no lines yielded more than the MG III and IV cultivars IA3010 or HS93-4118 (Table 1). Broad-sense heritability estimates calculated on a line-mean basis for all traits were moderate to high (Table 1) and were similar to those reported previously in other *G. max* QTL mapping populations (Mansur et al., 1993a; Orf et al., 1999; Specht et al., 2001).

Mean squares were estimated for the 167 F_5–derived lines grouped into sets according to maturity and grown across 12 environments. Mean square estimates for environment, sets, lines within sets, environment by set, and environment by lines within set interactions were significant ($P < 0.01$) for all traits measured except for protein concentration and seed yield where sets were not significant. Variance component estimates revealed which components accounted for the variance of each trait measured (Table 2). For yield, the environment accounted for more of the variance compared with lines within sets, environment by sets, and environment by lines within set interactions.

Agronomic and seed composition traits generally were not significantly ($P > 0.05$) correlated with yield (Table 3). Only protein concentration was significantly correlated with yield when data from all sets were included and this negative correlation was small. Days to maturity was positively correlated with yield within each set, but these correlations were relatively small, and yield and maturity were not correlated in the combined

Table 2. Variance component estimates for the F_5–derived lines of the BSR 101 × LG82-8379 soybean population grouped into sets and evaluated in 12 environments.

Variance component†	Days to flower (R1)	Days to maturity (R8)	Reproductive period (R8–R1)	Plant height	Lodging score (1–5)‡	Yield	Protein conc.	Oil conc.
				mm		kg ha^{-1}	g kg^{-1}	
Environment	1.6	278.8	296.1	31.8	0.66	215 308	1.05	0.10
Reps (environment)	0.1	0.7	0.2	2.7	0.04	5 302	0.11	0.03
Sets	18.6	102.5	31.0	35.4	0.05	0	0.00	0.03
Environment × sets	0.8	8.8	5.6	17.1	0.10	32 011	0.21	0.06
Lines (sets)	1.8	4.7	3.4	23.1	0.20	19 484	0.53	0.15
Environment × lines (sets)	1.5	1.5	2.7	10.8	0.08	12 971	0.13	0.02
Residual	3.0	5.1	7.5	38.5	0.30	106 655	0.95	0.26

† Variance components calculated by the REML option of VARCOMP (SAS Institute, Cary, NC).
‡ 1 = all plants standing erect, 5 = all plants prostrate.

Table 3. Pearson correlation coefficients between seed yield and other traits based on means measured in the BSR 101 × LG82-8379 soybean population of F_5–derived lines grouped within and across Sets 1, 2, and 3, according to maturity.

	Sets			
Trait†	1	2	3	Across
Days to flower (R1)	ns‡	ns	0.48***	ns
Days to maturity (R8)	0.31*	0.58***	0.34**	ns
Reproductive period (R8–R1)	0.28*	0.50***	ns	ns
Plant height, mm	ns	0.27*	ns	ns
Lodging score (1–5)§	ns	ns	ns	ns
Seed protein concentration, g kg^{-1}	ns	ns	ns	−0.18**
Seed oil concentration, g kg^{-1}	ns	ns	ns	ns

* Significant at the 0.05 probability level.
** Significant at the 0.01 probability level.
*** Significant at the 0.001 probability level.
† Days to maturity from planting, plant height, lodging, and seed yield averaged across 12 environments; days to flower from planting and reproductive period averaged across six environments; and seed protein and oil concentration averaged across eight environments.
‡ ns = nonsignificant.
§ 1 = all plants standing erect, 5 = all plants prostrate.

data. The lack of correlation between yield and the other measured traits indicate that this is a desirable population for identifying alleles that increase yield.

Marker and Linkage Analysis

Linkage analysis of the 145 polymorphic markers resulted in 26 linkage groups with 41 markers remaining unlinked. All linked and unlinked markers were assigned linkage group designations based on the USDA/Iowa State University composite genetic map (Cregan et al., 1999). Making the assumption that a QTL within 20 cM of either end of a linkage group or unlinked marker can be detected, 2454 cM or 82% of the soybean genome, estimated to be approximately 3000 cM, was analyzed for QTL (Shoemaker and Olson, 1993). Gaps of 20 cM or more existed between pairs of markers in this population. The maximum distance separating two markers was approximately 101 cM, which occurred in linkage group E.

QTL Analysis

In this study, we report QTL that were significantly associated ($P < 0.05$, LOD > 2.5) with agronomic and seed composition traits across 12 environments. Fifteen QTL were found significantly associated with seed yield (Table 4). At nine of the 15 QTL, the LG82-8379 alleles increased seed yield (Table 4A). Two QTL, located on linkage groups C2 and O, were found significant ($P < 0.05$) in combined analysis of all lines across the three sets. The alleles contributing to the increase in yield at these loci were both introgressed from FC 04007B. The linkage group C2 QTL explained 10% of the phenotypic variation and increased seed yield 60 kg ha^{-1}, while the linkage group O QTL explained 14% of the phenotypic variation and increased seed yield 47 kg ha^{-1}. Collectively, the two QTL accounted for 18% of the phenotypic variation for seed yield. The remaining seven QTL with the LG82-8379 alleles increasing seed yield were detected only within specific sets. The alleles contributing to the increase in yield on linkage groups A1, G, K, D2, and M were from PI 68508, whereas FC 04007B contributed the alleles on linkage groups B2 and H, in addition to alleles on linkage groups C2 and O. Within specific sets, these QTL explained 14 to 38% of the phenotypic variation and increased seed yield 60 to 148 kg ha^{-1}. Including the QTL that were significant across sets, the QTL collectively explained 47% of the pheno-

Table 4. Quantitative trait loci significantly ($P < 0.05$; LOD > 2.5) associated with seed yield in the BSR 101 × LG82-8379 soybean population. Allelic means based on seed yield are averaged across 12 environments.

						Allelic means			
Set†	LG‡	Marker	*P* value§	LOD¶	R^2#	BSR 101	LG82-8379	Effect††	Increase‡‡
						kg ha^{-1}			%
				A. Alleles from LG82-8379 increasing seed yield					
Across	C2	Satt363	<0.0001	–	0.10	2782	2842§§	60	2.2
Across	O	Satt358	0.0003	3.0	0.14	2802	2849§§	47	1.7
1	A1	Satt225	0.0050	2.6	0.14	2829	2889##	60	2.1
1	G	Satt394	0.0003	6.0	0.28	2815	2882##	67	2.4
1	K	Satt544	<0.0001	5.9	0.38	2802	2916##	114	4.1
2	B2	Satt168	0.0034	6.7	0.16	2775	2842§§	67	2.4
2	D2	Satt186	0.0272	3.2	0.21	2775	2842##	67	2.4
3	H	Satt142	<0.0001	2.7	0.35	2748	2896§§	148	5.4
3	M	Satt308	<0.0001	–	0.18	2748	2835##	87	3.2
				B. Alleles from BSR 101 increasing seed yield					
1	A1	Satt382	<0.0001	–	0.12	2903	2802	101	3.5
1	G	Satt191	<0.0001	–	0.27	2896	2815	81	2.8
1	N	Satt387	<0.0001	3.1	0.15	2903	2775	128	4.4
2	N	Satt339	0.0070	3.2	0.12	2849	2782	67	2.4
3	D1a	Satt547	0.0001	–	0.08	2842	2768	74	2.6
3	I	Satt440	<0.0001	–	0.15	2876	2762	114	4.0

† F_5–derived lines grouped within and across Sets 1, 2 and 3, according to maturity.
‡ LG, linkage group designation based on the USDA/Iowa State University composite genetic map.
§ *P* value based on pairwise comparison of least square means.
¶ LOD, composite interval mapping likelihood of odds; a dash indicates the marker was analyzed only by single marker-trait analysis.
# Phenotypic variance explained by individual marker.
†† Effect of positive allele on seed yield.
‡‡ Percentage increase in seed yield.
§§ The allele contributing to an increase in yield from FC 04007B.
The allele contributing to an increase in yield from PI 68508.

typic variation for seed yield within Set 1, 21% within Set 2, and 52% within Set 3.

Alleles from BSR 101 increased seed yield at six of the 15 yield QTL identified (Table 4B). The six QTL were detected only within specific sets. The QTL located on linkage groups A1, G, and N were detected within Set 1, N within Set 2, and D1a and I within Set 3. These QTL each explained 8 to 27% of the phenotypic variation and increased seed yield 67 to 128 kg ha^{-1}. The yield QTL in Set 1 lines located on linkage group N was approximately 32 cM away from the yield QTL in Set 2 lines on the same linkage group and was considered distinct. The QTL collectively accounted for 39% of the phenotypic variation for seed yield within Set 1, 12% within Set 2, and 15% within Set 3.

No QTL were found significant for R8 in combined analysis of all lines across the three sets, but four QTL were detected among lines of Set 1 (Table 5). The QTL located on linkage groups A1, G, K, and N, explained 6 to 12% of the phenotypic variation for R8 and accounted for a 1 to 2 d difference between the BSR 101 and LG82-8379 alleles at these loci. Collectively, the four QTL explained 24% of the phenotypic variation for R8. Twelve QTL were detected for plant height, one was significant across sets, four each within Sets 1 and 3, and three within Set 2 (Table 5). Each QTL explained 2 to 21% of the phenotypic variation for plant height and accounted for 20- to 60-mm differences between the BSR 101 and LG82-8379 alleles at these loci. The plant height QTL, including the QTL that was significant across sets, collectively accounted for 30% of the phenotypic variation within Set 1, 42% within Set 2, and 55% within Set 3. One QTL was detected for lodging within Set 3 on linkage group D1b. This QTL explained 6% of the phenotypic variation (Table 5) and the LG82-8379 allele at this locus increased lodging slightly. No QTL for R1 or reproductive period were identified.

Eleven QTL were significantly ($P < 0.05$, LOD > 2.5) associated with protein concentration (Table 5). Five QTL, located on linkage groups A2, C1, C2, D1b, and O, were significant in combined analysis of all lines across the three sets. Each QTL explained 5 to 16% of

Table 5. Quantitative trait loci significantly ($P < 0.05$; LOD > 2.5) associated with agronomic and seed composition traits in the BSR 101 × LG82-8379 soybean population.

						Allelic means††	
Set†	LG‡	Marker	*P* value§	LOD¶	R^2#	BSR 101	LG82-8379
			Days to maturity (R8)				
1	A1	Satt382	<0.0001	–	0.12	111	109
1	G	Satt394	<0.0001	2.6	0.06	109	111
1	K	Satt544	<0.0001	3.0	0.07	110	111
1	N	Satt387	<0.0001	2.5	0.11	111	109
			Plant height, mm				
Across	F	Satt490	<0.0001	4.4	0.08	950	910
1	A1	Satt382	<0.0001	–	0.05	870	850
1	G	Satt191	<0.0001	–	0.09	870	840
1	K	Satt544	<0.0001	3.6	0.06	840	870
1	N	Satt387	<0.0001	4.6	0.05	870	840
2	C2	Satt277	<0.0001	6.3	0.20	910	970
2	N	Satt339	0.0046	2.5	0.02	950	930
2	O	Satt478	<0.0001	3.6	0.11	960	930
3	C2	Satt277	<0.0001	7.1	0.21	950	990
3	D1a	Satt547	0.0002	–	0.07	960	980
3	D1b	Satt542	<0.0001	2.8	0.20	950	990
3	H	Satt142	0.0002	2.5	0.04	980	960
			Lodging score (1–5)‡‡				
3	D1b	Satt542	<0.0001	2.6	0.06	2.5	2.8
			Seed protein concentration, g kg^{-1}				
Across	A2	Satt409	<0.0001	3.6	0.06	424	428
Across	C1	Satt338	<0.0001	–	0.16	423	428
Across	C2	Satt363	<0.0001	–	0.05	428	425
Across	D1b	Satt157	<0.0001	4.3	0.14	428	424
Across	O	Satt358	<0.0001	3.0	0.09	424	427
1	F	Satt510	<0.0001	–	0.16	421	426
1	K	Satt544	0.0003	3.7	0.03	425	423
2	B2	Satt168	0.0009	2.7	0.10	424	426
2	N	Satt339	<0.0001	4.4	0.13	422	427
3	H	Satt142	0.0004	2.8	0.03	430	427
3	M	Satt308	<0.0001	–	0.10	427	430
			Seed oil concentration, g kg^{-1}				
Across	C1	Satt338	<0.0001	–	0.05	220	218
Across	D1b	Satt157	<0.0001	5.1	0.10	217	220
1	J	Satt285	<0.0001	–	0.16	219	222

† F_5–derived lines grouped within and across Sets 1, 2, and 3, according to maturity.
‡ LG, linkage group designation based on the USDA/Iowa State University composite genetic map.
§ *P* value based on pairwise comparison of least square means.
¶ LOD, composite interval mapping likelihood of odds; a dash indicates marker analyzed only by single marker-trait analysis.
# Phenotypic variance explained by individual marker.
†† Allelic means based on data averaged across environments.
‡‡ 1 = all plants standing erect, 5 = all plants prostrate.

the phenotypic variation for protein concentration and collectively accounted for 30%. The six remaining QTL associated with protein concentration were found only within specific sets. The QTL located on linkage groups B2, F, H, K, M, and N, each explained 3 to 16% of the phenotypic variation for protein concentration. Including the QTL that were significant across sets, the QTL collectively explained 50% of the phenotypic variation for protein concentration within Set 1, and 53% each within Sets 2 and 3. The difference in protein concentration between the BSR 101 and LG82-8379 alleles at each QTL ranged from 2 to 5 g kg^{-1}.

Three QTL were significantly ($P < 0.05$, LOD > 2.5) associated with oil concentration (Table 5). Two QTL, located on linkage group C1 and D1b, were significant in combined analysis of all lines across the three sets and explained 5 to 10% of the phenotypic variation for oil concentration. Collectively, the two QTL accounted for 10% of the phenotypic variation for oil concentration. The third QTL located on linkage group J was detected only within Set 1 and explained 16% of the phenotypic variation for oil concentration. Including the QTL that were significant across sets, the oil QTL collectively explained 26% of the phenotypic variation for oil concentration within Set 1. The difference in oil concentration between the BSR 101 and LG82-8379 alleles at these loci ranged from 2 to 3 g kg^{-1}.

DISCUSSION

There were 46 putative QTL significantly ($P < 0.05$, LOD > 2.5) associated with agronomic and seed composition traits across 12 environments. In this study, permissive detection limits ($P < 0.05$, LOD > 2.5) were used as we were less concerned with finding false positive QTL (Type I error) than we were in missing real QTL (Type II error). The experiment-wide error rate in this study is close to 100% and thus at least some of the QTL must be presumed to be false-positives. Therefore, the validity of each QTL identified will need to be confirmed. Gaps of 20 cM or more existed between pairs of markers; therefore, additional QTL for these traits are likely to exist.

To determine the impact of the significant environment by set and environment by lines within set interactions had on our QTL analysis for yield, we examined the effects of the 15 yield QTL in their respective sets in each of the 12 environments. The effect of the yield QTL was positive and statistically significant in 21% of the 180 (i.e., 15 × 12) yield QTL and environment combinations. In 60% of the yield QTL and environment combinations, the QTL effect was positive but not statistically significant. In only five of the yield QTL and environment combinations were the effects on yield negative, and none of these negative effects were statistically significant. Three of the five negative effects were associated with Satt142 on linkage group H. Although this QTL seems to be highly variable, the mean increase in yield among Set 3 lines associated with this QTL was the largest recorded for any of the QTL identified in this research.

Comparison of our results with those of others (Orf et al., 1999; Specht et al., 2001; Yuan et al., 2002) revealed QTL mapping to similar positions. Specht et al. (2001) identified a QTL, marked by Satt277, near our QTL on linkage group C2 that accounted for phenotypic variances of 13 to 15% for maturity, 5 to 15% for plant height, 6% for lodging, and 7 to 13% for seed yield in a RIL population derived from crossing Minsoy and Noir 1. In their study, the impact on seed yield from the beneficial Minsoy allele was attributed to the maturity locus *E*1 that is located 4 cM below Satt277 on linkage group C2 (Cregan et al., 1999). Orf et al. (1999) also discovered a QTL on linkage group C2 in a Noir 1 × Archer RIL population where the Noir 1 allele, marked by Satt277, increased seed yield 172 kg ha^{-1} and accounted for 11% of the phenotypic variance. In this case, Satt277 was not significantly associated with maturity. In our population, the LG82-8379 allele for Satt363 on linkage group C2 was associated with increased seed yield of 60 kg ha^{-1} and accounted for 10% of the phenotypic variance for yield across all sets. The LG82-8379 alleles in this region also were associated with increased plant height of 60 mm among Set 2 lines (R^2 = 20%) and 40 mm among Set 3 lines (R^2 = 21%) but were not associated with lodging. While Satt363 maps 15.2 cM from the maturity locus *E*1 on linkage group C2 (Cregan et al., 1999), no maturity differences were associated with Satt363 in our study.

Specht et al. (2001) identified a QTL, marked by Satt317, near our QTL on linkage group H, that accounted for phenotypic variances of 2% for plant height, 5% for lodging, and 3% for seed yield in a Minsoy × Noir 1 RIL population. The QTL association in their population was attributed to the Noir 1 semisparse pubescence (*Ps-s*) allele located near Satt317. Lines homozygous for the Noir 1 *Ps-s* allele tended to be spindly, lodging-prone, and generally yielded less than lines homozygous for the Minsoy *ps* allele. In Set 3 lines of our population, the LG82-8379 allele on linkage group H, marked by Satt142, increased seed yield 148 kg ha^{-1} (R^2 = 35%) and decreased plant height 20 mm (R^2 = 4%). This genetic region was not associated with lodging and there was no obvious segregation for pubescence density in our population. On the University of Utah genetic linkage map (Cregan et al., 1999), Satt317 is 4.2 cM from Satt142 on linkage group H.

Yuan et al. (2002) identified yield QTL on linkage group K in two RIL populations. 'Essex' provided the beneficial allele (R^2 = 10%, 110 kg ha^{-1} yield increase) marked by Satt337 in their Essex × 'Forrest' RIL population, whereas 'Flyer' provided the beneficial allele (R^2 = 15%, 220 kg ha^{-1} yield increase) marked by Satt326 in their Flyer × 'Hartwig' RIL population. A QTL on linkage group K, marked by Satt326, also was identified by Specht et al. (2001) in a Minsoy × Noir 1 RIL population and it accounted for phenotypic variances of 2% for maturity and 2% for seed yield in only 1 yr of a 2 yr study. In our Set 1 lines, Satt544 marked the LG82-8379 allele on linkage group K that was associated with increasing seed yield 114 kg ha^{-1} (R^2 = 38%), increasing days to maturity by 1 d (R^2 = 7%), and

increasing plant height by 30 mm (R^2 = 6%). Markers Satt326 and Satt337 map approximately 11.3 cM from Satt544 on linkage group K of the USDA/Iowa State University genetic linkage map (Cregan et al., 1999).

Specht et al. (2001) reported that 4% of the phenotypic variance for seed yield within a Minsoy × Noir 1 RIL population was accounted for by segregation at the Rpg_4 locus on linkage group N for resistance/susceptibility to bacterial blight. Bacterial blight occurred within the test the year this association was made and lines homozygous for the susceptible Noir 1 allele yielded 181 kg ha^{-1} less. In our population, the allele from BSR 101 on linkage group N, marked by Satt387 (R^2 = 15%) increased seed yield 128 kg ha^{-1} among Set 1 lines while Satt339 (R^2 = 12%) increased seed yield 67 kg ha^{-1} among Set 2 lines. The BSR 101 allele at Satt387 also was associated with increasing days to maturity by 2 d (R^2 = 11%) and increasing plant height 30 mm (R^2 = 5%). At Satt339, the BSR 101 allele also was associated with increasing plant height 20 mm (R^2 = 2%). On the University of Utah genetic linkage map (Cregan et al., 1999), the Rpg_4 locus on linkage group N is 7.5 cM from Satt387 and 23 cM from Satt339 that marks the QTL in our population.

Several yield QTL identified in this population also were associated with other agronomic traits. For example, the marker alleles associated with an increase in seed yield on linkage groups A1, G, K, and N also were associated with an increase in days to maturity and with taller plants. In contrast, marker alleles that were associated with greater yield on linkage groups H and D1a were associated with shorter plants. Although maturity and plant height cosegregated with several of the yield QTL, the variation observed was small and it is uncertain whether these differences would account for the change in seed yield associated with these regions. The correlation coefficients between seed yield and these traits were generally small and even though both plant height and maturity were significantly different between sets, seed yield was not. Further genetic dissection of the regions containing these QTL would be needed to distinguish between pleiotropy or gene linkage to define these relationships.

Identifying QTL with positive effects on seed yield while simultaneously improving or at least not decreasing protein concentration would be beneficial. Eight seed yield QTL and four protein concentration QTL were found to segregate independently of each other; however, three of the 15 yield QTL identified also were associated with increasing protein concentration. The yield-increasing LG82-8379 alleles, located on linkage groups B2, M, and O, also were associated with increased protein concentration of 2 to 3 g kg^{-1}. Unique seed yield and protein QTL may be segregating within this population.

Despite the observation of significant variation among lines within each set of this population and the high heritability estimates calculated for each trait measured (Table 1), many QTL identified were found only within lines of specific sets. It is not clear whether this is the result of a particular QTL identified only having an effect in certain maturities or whether there was insufficient resolution to detect a QTL within a particular set. The number of lines evaluated within each set was low, which can limit precision of QTL localization, cause an underestimation of the number of QTL controlling the trait, overestimate the effects of QTL, and reduce the power of QTL detection (Beavis, 1998). Confirmation studies will determine if the QTL identified in this study, within and among sets, are truly real or not. Three set-specific yield QTL that we identified in our population were found by other researchers to be significant within lines of different maturities in their populations (Specht et al., 2001; Yuan et al., 2002). This indicates that these particular yield QTL may not be specific to maturity groups or may be specific to maturity groups within certain genetic backgrounds. We have also determined that the early MG IV experimental line, LG91-7350R, previously developed from BSR 101 × LG82-8379 and used as a check in this study, carries the yield enhancing, Set 2 QTL marked by Satt168 on linkage group B2 from FC 04007B. Further examination of the environmental or developmental conditions the set-specific QTL respond to will be necessary.

Results from this study indicate that soybean PIs contain QTL alleles that may be of benefit to North American soybean cultivars. Of particular interest is the yield QTL marked by Satt358 on linkage group O. The QTL was significant in the combined analysis across sets of this population. For this QTL, the FC 04007B allele enhanced yield 47 kg ha^{-1} and increased seed protein concentration 3 g kg^{-1}. To date, no other studies have detected a yield QTL at this chromosomal location. Confirmation of the yield-enhancing QTL from LG82-8379 is currently being performed in the genetic background of the population used in the current study and in different genetic backgrounds.

ACKNOWLEDGMENTS

This research was supported by a grant from the Illinois Soybean Program Operating Board.

REFERENCES

Apuya, N., B. Frazier, P. Keim, E.J. Roth, and K.G. Lark. 1988. Restriction length polymorphisms as genetic markers in soybean, *Glycine max* (L.) Merr. Theor. Appl. Genet. 75:889–901.

Beavis, W.D. 1998. QTL analyses: Power, precision, and accuracy. p. 145–162. *In* A.H. Paterson (ed.) Molecular dissection of complex traits. CRC Press, New York.

Bernard, R.L., G.A. Juvik, E.E. Hartwig, and C.J. Edwards, Jr. 1988. Origins and pedigrees of public soybean varieties in the United States and Canada. USDA Technical Bulletin No. 1746. U.S. Gov. Print. Office, Washington, DC.

Cianzio, S.R., S.P. Schultz, W.R. Fehr, and H. Tachibana. 1991. Registration of 'Archer' soybean. Crop Sci. 31:1707.

Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. VanToai, D.G. Lohnes, J. Chung, and J.E. Specht. 1999. An integrated genetic linkage map of the soybean. Crop Sci. 39:1464–1490.

Cregan, P.B., and C. Quigley. 1997. Simple sequence repeat DNA marker analysis. p. 173–185. *In* G. Caetano-Anolles and P.M. Gresshoff (ed.) DNA markers: Protocols, applications and overview. J. Wiley and Sons, New York.

Fehr, W.R. 1987. Heritability. p. 95–105. *In* W.R. Fehr (ed.) Principles of cultivar development. Volume I, Theory and technique. Macmillan, New York.

Fehr, W.R., C.E. Caviness, D.T. Burmood, and J.S. Pennington. 1971. Stage of development descriptions for soybeans, *Glycine Max* (L.) Merrill. Crop Sci. 11:929–933.

Gizlice, A., T.E. Carter, Jr., and J.W. Burton. 1993. Genetic diversity in North American Soybean: I. Multivariate analysis of founding stock and relation to coefficient of parentage. Crop Sci. 33:614–620.

Gizlice, Z., T.E. Carter, Jr., and J.W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci. 34(5):1143–1151.

Keim, P., T.C. Olson, and R.C. Shoemaker. 1988. A rapid protocol for isolating soybean DNA. Soybean Genet. Newsl. 15:150–152.

Keim, P., R.C. Shoemaker, and R.G. Palmer. 1989. Restriction fragment length polymorphism diversity in soybean. Theor. Appl. Genet. 77:786–792.

Kisha, T.J., C.H. Sneller, and B.W. Diers. 1997. Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop Sci. 37:1317–1325.

Lander, E.S., and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199.

Mansur, L.M., K.G. Lark, H. Kross, and A. Oliveira. 1993a. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). Theor. Appl. Genet. 86:907–913.

Mansur, L.M., J.H. Orf, and K.G. Lark. 1993b. Determining linkage of quantitative trait loci to RFLP markers using extreme phenotypes of recombinant inbreds of soybean [*Glycine max* (L.) Merr.]. Theor. Appl. Genet. 86:914–918.

McIntosh, M.S. 1983. Analysis of combined experiments. Agron. J. 75:153–155.

Orf, J.H., K. Chase, T. Jarvik, L.M. Mansur, P.B. Cregan, F.R. Adler, and K.G. Lark. 1999. Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. Crop Sci. 39:1642–1651.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular cloning, a laboratory manual. 2nd ed. Cold Spring Harbor Laboratory Press, Plainview, NY.

Satterthwaite, F.E. 1946. An approximate distribution of estimates of variance components. Biom. Bull. 2:110–114.

Shoemaker, R.C., and T.C. Olson. 1993. Molecular linkage map of soybean (*Glycine max* L. Merr.). p. 6.131–6.138. *In* S.J. O'Brien (ed.) Genetic maps: Locus maps of complex genomes. Cold Spring Harbor Laboratory Press, Plainview, NY.

Specht, J.E., K. Chase, M. Macrander, G.L. Graef, J. Chung, J.P. Markwell, M. Germann, J.H. Orf, and K.G. Lark. 2001. Soybean response to water: A QTL analysis of drought tolerance. Crop Sci. 41:493–509.

Tanksley, S.D., and S.R. McCouch. 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277: 1063–1066.

Terry, L.I., K. Chase, T. Jarvik, J. Orf, L. Mansur, and K.G. Lark. 2000. Soybean quantitative trait loci for resistance to insects. Crop Sci. 40(2):375–382.

Thompson, J.A., P.J. Amdor, and R.L. Nelson. 1999. Registration of LG90-2550 and LG91-7350R Soybean Germplasm. Crop Sci. 39: 302–303.

Thompson, J.A., and R.L. Nelson. 1998. Utilization of diverse germplasm for improving soybean yield. Crop Sci. 38:1362–1368.

Thompson, J.A., R.L. Nelson, and L.O. Vodkin. 1998. Identification of diverse soybean germplasm using RAPD markers. Crop Sci. 38:1348–1355.

Utz, H.F., and A.E. Melchinger. 1996. PLABQTL: A program for composite interval mapping of QTL. J. Quant. Trait Loci 2:239–243. No longer available; see http://www.unihohenheim.de/~ipspwww/soft.html (verified 27 Jan. 2004).

Van Ooijen, J.W., and R.E. Voorrips. 2001. JoinMap® Version 3.0: Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, the Netherlands.

VanToai, T.T., S.K. St-Martin, K. Chase, G. Boru, V. Schnipke, A.F. Schmitthenner, and K.G. Lark. 2001. Identification of a QTL associated with tolerance of soybean to soil waterlogging. Crop Sci. 41(4):1247–1252.

Wang, D., J. Shi, S.R. Carlson, P.B. Cregan, R.W. Ward, and B.W. Diers. 2003. A low-cost, high-throughput polyacrylamide gel electrophoresis system for genotyping with microsatellite DNA markers. Crop Sci. 43:1828–1832.

Yuan, J., V.N. Njiti, K. Melsem, J.J. Iqbal, K. Triwitayakorn, M.A. Kassem, G.T. Davis, M.E. Schmidt, and D.A. Lightfoot. 2002. Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance. Crop Sci. 42:271–277.